1 月 January
2019, 62(1): 101 – 116

昆 虫 学 报
ACTA ENTOMOLOGICA SINICA

http: // www. insect. org. cn
doi: 10. 16380/j. kcxb. 2019. 01. 011

Sequencing of the complete mitochondrial genome of *Anopheles lindesayi* and a phylogenetic analysis of the genus *Anopheles* (Diptera: Culicidae) based on mitochondrial genomes

MAO Qi-Meng, LI Ting-Jing, FU Wen-Bo, YAN Zhen-Tian, CHEN Bin*

(Chongqing Key Laboratory of Vector Insects, Institute of Entomology and Molecular Biology,
Chongqing Normal University, Chongqing 401331, China)

Abstract: 【Aim】 To sequence and analyze the complete mitochondrial genome (mtgenome) of *Anopheles lindesayi*, and to construct and discuss the molecular phylogenetic relationships of the genus *Anopheles* based on the known mtgenomes. 【Methods】 The complete mtgenome of *An. lindesayi* was sequenced and annotated, and its general features and base composition were analyzed. The phylogenetic relationships between *An. lindesayi* and other 32 species in the genus *Anopheles* were reconstructed using maximum likelihood (ML) and Bayesian inference (BI) methods based on the concatenated nucleotide sequences and amino acid sequences of 13 protein-coding genes (PCGs), and the systematics of the genus *Anopheles* was discussed based on the phylogenetic analysis. 【Results】 The complete mtgenome sequence of *An. lindesayi* is 15 366 bp in length, which contains 13 PCGs, 22 tRNA genes, two rRNA genes, and one control region. The mtgenome has a clear bias in nucleotide composition with a positive AT-skew and a negative GC-skew. The initiation codons of PCGs comply with the ATN rule except that *COX*1 uses TCG and *ND*5 uses GTG as the start codon, and the termination codon is TAA or incomplete T. The tRNAs have the typical clover-leaf structure, but tRNA$^{Ser(AGN)}$ has a large loop instead of the conserved stem-and-loop structure. The control region has the highest AT content (94.54%). The sliding window analysis suggested that the PCGs are the most suitable markers to elucidate the phylogenetic relationships at the subgenus and genus levels. The phylogenetic analysis strongly supported each of the subgenera *Cellia*, *Anopheles*, *Nyssorhynchus* and *Kerteszia* to be a monophyletic group. *An. lindesayi* and *An. atroparvus* + *An. quadrimaculatus* A, the latter two traditionally classified as *Anopheles* Series, are separated by *An. sinensis* in Myzorhynchus Series on all the four phylogenetic trees, raising a new debate on the taxonomy of the two Series. 【Conclusion】 This study obtained the complete mtgenome sequence of *An. lindesayi* and explored the mtgenome characteristics and phylogeny of the genus *Anopheles*, establishing the information frame for further studying the mtgenomes and phylogeny of the family Culicidae.

Key words: Culicidae; *Anopheles*; *Anopheles lindesayi*; mitochondrial genome; nucleotide composition; phylogeny

1 INTRODUCTION

In the recent years, mitochondrial genomes (mtgenomes) as ideal molecular markers have gained importance for comprehensive evolutionary and population genetic studies of insects, due to their relatively small genome sizes, low levels of recombination and maternal inheritance (Wei *et al.*, 2010; Zhao *et al.*, 2010). Arrangement of genes in mtgenomes is usually stable, retaining the ancestral pattern of gene arrangement (Cameron, 2014; Hao *et al.*, 2017). For most Diptera insects, mtgenomes are 14 – 20 kb in size, and all of them consist of 13 protein coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, and a control region (CR, also called the AT-rich region) on a double-stranded circular DNA molecule (Boore, 1999; Hua *et al.*, 2016; Oliveira *et al.*, 2016; Hao *et al.*, 2017).

The genus *Anopheles* of mosquitoes is distributed in tropical and temperate regions of the world (Harbach and Kitching, 2016), and many species

---

基金项目: 国家自然科学基金项目(31872262, 31672363); 国家科技基础性工作专项重点项目(2015FY210300); 广西科技重大专项(桂科 AA17129002)

作者简介: 毛启萌, 女, 1993 年 3 月生, 四川乐山人, 硕士研究生, 研究方向为生物化学与分子生物学, E-mail: 874561386@ qq. com

* 通讯作者 Corresponding author, E-mail: bin. chen@ cqnu. edu. cn

收稿日期 Received: 2018-03-15; 接受日期 Accepted: 2018-10-08

of the genus are important vectors of malaria. The phylogenetic relationships in the genus remain unsettled, despite some phylogenetic studies have earlier been conducted based on morphological characteristics and individual genes of molecular datasets (Foley *et al.*, 1998; Sallum *et al.*, 2000; Harbach and Kitching, 2005, 2010). With the development of molecular biology techniques, whole genome sequencing is being widely used for the studies on evolutionary biology, and the mtgenome sequences can be more phylogenetically informative than shorter sequences of individual genes (Boore *et al.*, 2005; Boore, 2006; Masta and Boore, 2008).

Anopheles lindesayi belongs to the subgenus *Anopheles*, Anopheles Series, Lindesayi Group. It is widely distributed from India, Afghanistan and Pakistan, eastward to mainland China, Korea, Japan, Taiwan province of China, and southeastward to the Philippines, Thailand and Malaysia (Reid, 1968; Harrison *et al.*, 1990; Glick, 1992; Sames *et al.*, 2008). Its larvae inhabit in stream pools, rock holes or cavities containing fresh water, and its adults perch in the bullpen (Sames *et al.*, 2008; Kim *et al.*, 2009). The cytochrome-oxidase subunit 1 gene (*COX*1) of mtgenome and the internal transcribed spacer 2 (ITS2) of *An. lindesayi* have been applied for the molecular identification, phylogenetic and population genetics analysis involving in the species (Wang *et al.*, 2012; Ying *et al.*, 2013), whereas the complete mtgenome sequence of the species has not been reported.

In this study, we sequenced the complete mtgenome of *An. lindesayi*, and analyzed its characteristics including the composition and biases of nucleotides, codon usage, tRNA secondary structure, the CR and nucleotide diversity. Importantly, to discuss the characteristics and evolution of mtgenomes in the genus *Anopheles*, we carried out the comparative mtgenome analysis of *An. lindesayi* and other 32 species in the genus *Anopheles*, reconstructed the phylogenetic relationships using concatenated nucleotide and amino acid (AA) sequences of 13 PCGs of these mtgenome sequences. This study obtained the complete mtgenome sequence of *An. lindesayi* and explored the mtgenome characteristics and phylogeny of the genus *Anopheles*, and further established the information frame for the mtgenome and phylogenetic study of the family Culicidae.

2 MATERIALS AND METHODS

2.1 Sample collection and DNA extraction

The specimens of *An. lindesayi* were collected on June 20, 2014 from field in Shanxi Province, China (31.22°N, 108.44°E, altitude: 673 m). The mosquitoes were identified by their morphological characteristics, and further determined as *An. lindesayi* by sequencing *COX*2 (GenBank no.: JX070689.1), which has 99% nucleotide sequence identity with the known sequences of *An. lindesayi* download from the NCBI database (https://www.ncbi.nlm.nih.gov/). Individual specimens were preserved in 100% ethyl alcohol and stored at −80℃ in Institute of Entomology and Molecular Biology, Chongqing Normal University for long-term preservation. The total DNA from a female mosquito was isolated with the TIANamp Genomic DNA Kit (TianGen, China) for mtgenome sequencing in the present study.

2.2 PCR amplification and sequencing of the *An. lindesayi* mtgenome

The complete mtgenome of *An. lindesayi* was amplified using 18 PCR primer pairs designed in Chongqing Normal University (Zhang *et al.*, 2013). A 25 μL reaction of PCR was used containing 4 μL of dNTPs, 1 μL of each primer (10 μmol/L), 2.5 μL of 10 × LA PCR Buffer I, 3 μL of DNA templates, 0.25 μL of LA Taq polymerase (TaKaRa, Dalian) and 13.25 μL ddH_2O. The PCR performed by initial denaturation at 94℃ for 1 min, 34 – 35 cycles of denaturation at 94℃ for 40 s, annealing at 48 – 52℃ for 45 s, extension at 68℃ for 1 min and final elongation at 72℃ for 10 min. The PCR products were verified by electrophoresis on a 1% agarose gel, and sent to the Sangon Biotech (Shanghai) Company Limited for sequencing if the amplified fragments were clear in electropherograms. All fragments were sequenced in both directions.

2.3 Sequence assembly, annotation and analysis

The complete mtgenome sequences of *An. lindesayi* were assembled using DNAMAN v. 4.0 (Lynnon Biosoft). Their PCGs and rRNAs were identified by comparing with other mosquito mtgenome sequences (downloaded from the GenBank) using BioEdit (Alzohairy, 2011). Twenty tRNA genes were identified, and the stem-loop secondary structures were predicted using the tRNAscan-SE Search Server v. 1.21 (Lowe and Eddy, 1997), while the other two tRNAs ($tRNA^{Ser(AGN)}$ and $tRNA^{Arg}$) that could not be found by tRNAscan-SE Search Server v. 1.21 were identified in comparison with the known mosquito mtgenome sequences (downloaded from the GenBank). The AT-skew and GC-skew were calculated using the formulas AT skew = (A − T)/(A + T) and GC skew = (G − C)/(G + C), respectively (Perna and Kocher, 1995). The base composition, codon usage and the relative synonymous codon usage (RSCU) were calculated by MEGA 5 (Tamura *et al.*, 2014). The four mtgenome sequences of the *Anopheles* subgenus were aligned using BioEdit, and sliding window analysis was performed using DnaSP v. 5. 10. 01 (Librado and Rozas, 2009), to reveal the nucleotide diversity at different nucleotide positions throughout the whole mtgenomes.

2.4 Phylogenetic analysis

The phylogenetic relationships of 33 mtgenome sequences of *Anopheles* were analyzed using both maximum likelihood (ML) and Bayesian inference (BI) methods with the mtgenome sequence of *Culex pipiens pallens* (KT851543) as the outgroup (Table 1). Among all of the 33 mtgenome sequences, the mtgenome of *An. lindesayi* was newly sequenced in this study, that of *An. sinensis* has earlier been reported (Hao *et al.*, 2017) but not yet available in the GenBank, and all the other 31 mtgenome sequences were downloaded from the GenBank. The nucleotide and amino acid sequences of 13 protein-coding genes were separately aligned using MAFFT (Kazutaka and Standley, 2013). After removing the poorly aligned and divergent regions using Gblocks

Table 1 Accession numbers and information of mtgenomes used in the present phylogenetic analysis

Subgenus/species	GenBank accession number	Size (bp)	Genome without CR			References
			AT (%)	AT-skew	GC-skew	
Anopheles						
An. lindesayi	KX961140	15 366	78.3	0.029	-0.145	This study
An. atroparvus	NC028213	15 458	76.7	0.036	-0.147	Directly downloaded from GenBank
An. quadrimaculatus A	NC000875	15 455	77.4	0.040	-0.180	Mitchell *et al.*, 1993
An. sinensis	MF322628	15 418	78.4	0.030	-0.160	Unpublished
Cellia						
An. arabiensis	NC028212	15 369	77.0	0.035	-0.154	Directly downloaded from GenBank
An. christyi	NC028214	14 967	76.7	0.044	-0.173	Directly downloaded from GenBank
An. coluzzii	NC028215	15 441	77.0	-0.036	-0.153	Directly downloaded from GenBank
An. cracens	NC020768	15 412	77.4	0.029	-0.156	Logue *et al.*, 2013
An. culicifacies B	NC027502	15 330	78.0	0.033	-0.151	Hua *et al.*, 2016
An. culicifacies	NC028216	15 364	77.5	0.040	-0.160	Directly downloaded from GenBank
An. dirus A	JX219731	15 404	77.6	0.030	-0.160	Logue *et al.*, 2013
An. epiroticus	NC028217	15 379	77.2	0.030	-0.140	Directly downloaded from GenBank
An. farauti 1	JX219741	15 412	77.8	0.040	-0.150	Logue *et al.*, 2013
An. farauti 4	NC020770	15 412	77.4	0.030	-0.150	Logue *et al.*, 2013
An. gambiae	NC002084	15 363	77.0	0.030	-0.150	Beard *et al.*, 1993
An. hinesorum	NC020769	15 336	77.5	0.040	-0.150	Logue *et al.*, 2013
An. koliensis	JX219742	15 412	77.5	0.040	-0.160	Logue *et al.*, 2013
An. maculatus	NC028218	14 850	77.6	0.040	-0.150	Directly downloaded from GenBank
An. melas	NC028219	15 366	77.0	0.040	-0.160	Directly downloaded from GenBank
An. merus	NC028220	15 365	77.0	0.040	-0.160	Directly downloaded from GenBank
An. minimus	KT895423	15 395	78.6	0.026	-0.141	Hua *et al.*, 2016
An. punctulatus	NC028222	15 045	78.2	0.040	-0.130	Logue *et al.*, 2013
Nyssorhynchus						
An. albitarsis F	NC030718	15 418	77.3	0.031	-0.162	Krzywinski *et al.*, 2011
An. albitarsis G	NC030716	15 474	77.0	0.031	-0.162	Krzywinski *et al.*, 2011
An. albitarsis	NC020662	15 413	77.1	0.031	-0.162	Krzywinski *et al.*, 2011
An. darlingi	NC014275	15 386	77.6	0.030	-0.140	Moreno *et al.*, 2010
An. deaneorum	NC020663	15 424	77.2	0.030	-0.170	Krzywinski *et al.*, 2011
An. janconnae	NC030717	15 422	77.2	0.030	-0.160	Krzywinski *et al.*, 2011
An. oryzalimnetes	NC030715	15 425	77.1	0.030	-0.160	Krzywinski *et al.*, 2011
Kerteszia						
An. bellator	NC030249	15 668	77.5	0.028	-0.182	Oliveira *et al.*, 2016
An. laneanus	NC030250	15 446	77.7	0.021	-0.162	Oliveira *et al.*, 2016
An. cruzii	NC024740	15 449	78.4	0.018	-0.165	Directly downloaded from GenBank
An. homunculus	NC030248	15 739	77.7	0.020	-0.163	Oliveira *et al.*, 2016

(Talavera and Castresana, 2007), the individual alignments were then concatenated following their orders in the mitogenome. The best-fit substitution model for nucleotide datasets and AA datasets were selected by Modeltest using the Akaike Information Criterion (AIC) (Posada and Crandall, 1998) and ProtTest, and the GTR + I + G and WAG + G + I + F were determined as the best-fit models for nucleotide and AA sequences, respectively. RAxML (Stamatakis, 2014) was used in the ML analysis under the best-fit model detected, and the bootstrap support values were calculated based on 1 000 replicates. MrBayes (Ronquist and Huelsenbeck, 2003) was used in the BI analysis with the two independent runs and four chains (three heat and one cold) were performed simultaneously for 1 000 000 generations. The runs were stopped till the average deviation of split frequencies fall below 0.01, and the phylogenetic trees being sampled every 100 generations with 25% burn-in rate. The topology of the best-scoring trees were visualized and edited in MEGA 5. Bootstrap support values larger than 50%, and the names of four subgenera and eight Series in the genus *Anopheles* were marked on the corresponding branches of the constructed trees.

3 RESULTS

3.1 Structure and nucleotide composition of *An. lindesayi* mtgenome

The complete mtgenome of *An. lindesayi* is a double-stranded circular DNA molecule of 15 366 bp in size (GenBank accession number: KX961140) (Fig. 1). It contains 13 PCGs, 2 rRNAs, 22 tRNAs and a control region (CR, also called the AT-rich region) (Table 2). Twenty-two genes (nine PCGs and 13 tRNAs) are on the majority coding strand (J-strand), while the other 15 genes (four PCGs, nine tRNAs and two rRNAs) are on the minority strand (N-strand).

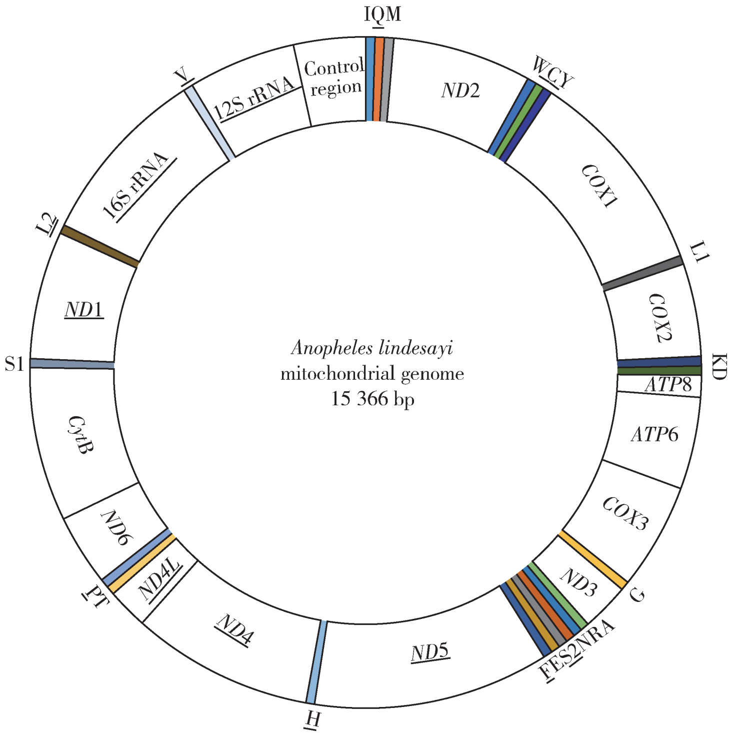

Fig. 1 Mitochondrial genome structure of *Anopheles lindesayi*

The unfilled white blocks denote protein-coding genes, ribosomal RNA (rRNA) and control regions (CR), while the color-filled blocks indicate transfer RNAs (tRNAs). Genes without underline are located on the majority strand, whereas those with underline are on the minority strand.

The mtgenome of *An. lindesayi* is slightly smaller than 25 completely sequenced *Anopheles* mtgenomes, but slightly bigger than seven completely sequenced *Anopheles* mtgenomes, whose sizes range from 14 850 bp (*An. maculatus*) to 15 739 bp (*An. homunculus*). In all the 33 mtgenoems investigated, the PCGs, tRNAs and rRNAs have minimal length variation in four *Anopheles* subgenera, whereas the CR has very large difference in length. There are seven intergenic spacers detected in the *An. lindesayi* mtgenome with a total length of 45 bp, ranging from 1 bp to 18 bp, with the longest spacer between $\text{tRNA}^{\text{Ser(UCN)}}$ and *ND*1. There are 14 genes to be overlapped with a total overlapping length of 33 bp, overlapping from 1 bp to 7 bp, and the longest overlapping are located between *ATP*8 and *ATP*6, and between *ND*4 and *ND4L* (Table 2).

Table 2 Structure of the *Anopheles lindesayi* mtgenome

Gene	Direction	Location (bp)	Size (bp)	Anti codon	Start codon	Stop codon	Intergenic nucleotides
tRNA^{Ile}	J	1 - 68	68	GAT			
tRNA^{Gln}	N	66 - 134	69	TTG			-3
tRNA^{Met}	J	134 - 202	69	CAT			-1
*ND*2	J	203 - 1 228	1 026		ATT	TAA	0
tRNA^{Trp}	J	1 227 - 1 295	69	TCA			-2
tRNA^{Cys}	N	1 295 - 1 358	64	GCA			-1
tRNA^{Tyr}	N	1 359 - 1 424	66	GTA			0
*COX*1	J	1 423 - 2 959	1 537		TCG	T	-2
$\text{tRNA}^{\text{Leu(UUR)}}$	J	2 960 - 3 025	66	TAA			0
*COX*2	J	3 027 - 3 711	685		ATG	T	1
tRNA^{Lys}	J	3 712 - 3 782	71	CTT			0
tRNA^{Asp}	J	3 795 - 3 862	68	GTC			12
ATP8	J	3 863 - 4 024	162		ATT	TAA	0
*ATP*6	J	4 018 - 4 698	681		ATG	TAA	-7
*COX*3	J	4 698 - 5 484	787		ATG	T	-1
tRNA^{Gly}	J	5 485 - 5 551	67	TCC			0
*ND*3	J	5 552 - 5 905	354		ATA	TAA	0
tRNA^{Arg}	J	5 904 - 5 967	64	TCG			-2
tRNA^{Ala}	J	5 968 - 6 033	66	TGC			0
tRNA^{Asn}	J	6 034 - 6 101	68	GTT			0
$\text{tRNA}^{\text{Ser(AGN)}}$	N	6 101 - 6 167	67	GCT			-1
tRNA^{Glu}	J	6 169 - 6 235	67	TTC			1
tRNA^{Phe}	N	6 234 - 6 300	67	GAA			-2
*ND*5	N	6 300 - 8 042	1 743		GTG	TAA	-1
tRNA^{His}	N	8 043 - 8 106	64	GTG			0
*ND*4	N	8 107 - 9 448	1 342		ATG	T	0
ND4L	N	9 442 - 9 741	300		ATG	TAA	-7
tRNA^{Thr}	J	9 747 - 9 811	65	TGT			5
tRNA^{Pro}	N	9 812 - 9 877	66	TGG			0
*ND*6	J	9 880 - 10 404	525		ATT	TAA	2
CytB	J	10 404 - 11 541	1 138		ATG	TAA	-1
$\text{tRNA}^{\text{Ser(UCN)}}$	J	11 540 - 11 605	66	TGA			-2
*ND*1	N	11 624 - 12 568	945		ATT	TAA	18
$\text{tRNA}^{\text{Leu(CUN)}}$	N	12 575 - 12 640	66	TAG			6
16S rRNA	N	12 641 - 13 967	1 327				0
tRNA^{Val}	N	13 968 - 14 039	72	TAC			0
12S rRNA	N	14 040 - 14 835	796				0
CR		14 836 - 15 366	531				0

J: J-strand; N: N-strand.

The mtgenome of *An. lindesayi* shows a high nucleotide bias with 78.87% of AT and 21.13% of GC (40.43% A, 38.44% T, 9.01% G, and 12.12% C). The AT contents of PCGs, rRNAs, tRNAs and CR are 77.44%, 82.05%, 79.00% and 94.54%, respectively. The whole mtgenome has positive AT-skew value (0.025) and negative GC-skew value (-0.148), and the different types of individual genes show different nucleotide biases on the J-strand and N-strand (Table 3).

Table 3 Nucleotide composition of the *Anopheles lindesayi* mtgenome

Feature	Proportion of nucleotides						
	T%	C%	A%	G%	A + T%	AT skew	GC skew
Whole genome	38.44	12.12	40.43	9.01	78.87	0.025	−0.148
Protein-coding genes	44.55	10.50	32.89	12.06	77.44	−0.151	0.069
1st codon position	38.83	9.88	32.62	18.67	71.45	−0.087	0.308
2nd codon position	46.28	18.81	20.53	14.39	66.80	−0.385	−0.133
3rd codon position	48.55	2.81	45.53	3.11	94.08	−0.032	0.050
Protein-coding genes on the J-strand	42.19	12.33	34.17	11.31	76.36	−0.105	−0.043
1st codon position	35.75	11.68	34.09	18.48	69.83	−0.024	0.225
2nd codon position	44.46	21.07	21.12	13.35	65.58	−0.356	−0.224
3rd codon position	46.38	4.23	47.29	2.09	93.67	0.010	−0.338
Protein-coding genes on the N-strand	48.31	7.59	30.86	13.24	79.17	−0.220	0.271
1st codon position	43.75	7.01	30.28	18.96	74.03	−0.182	0.460
2nd codon position	49.17	15.21	19.58	16.04	68.75	−0.430	0.027
3rd codon position	52.01	0.56	42.71	4.72	94.72	−0.098	0.789
tRNA genes	38.9	9.08	40.11	11.91	79.00	0.015	0.135
tRNA genes on the J-strand	38.14	10.31	41.01	10.54	79.15	0.036	0.011
tRNA genes on the N-strand	39.97	7.50	38.66	13.87	78.63	−0.017	0.298
rRNA	43.15	6.41	38.91	11.54	82.05	−0.052	0.286
Control region	50.28	3.95	44.26	1.51	94.54	−0.064	−0.448

AT skew = (A − T)/(A + T); GC skew = (G − C)/(G + C).

3.2 Characteristics and nucleotide comparison of *An. lindesayi* mtgenome

The 13 PCGs of *An. lindesayi* mtgenome have a total length of 11 225 bp, accounting for 73.05% of the whole length of the mtgenome. The PCGs have a higher frequency of T and G on the N-strand, whereas a higher frequency of T and C on the J-strand. In the PCGs of *An. lindesayi* mtgenome, the AT content of the 3rd codon position is obviously higher than those of the 1st and 2nd (Table 3). In the 13 PCGs, *COX*1 has the lowest AT content (70.8%) and *ND*6 has the highest AT content (85.7%) (Fig. 2: A). The AT-skew values range from −0.290 (*ND*1) to −0.037 (*COX*2), and the GC-skew values range from −0.333 (*ATP*8) to 0.306 (*ND4L*) (Fig. 2: B).

In the *An. lindesayi* mtgenome, 13 PCGs initiate with ATN as the start codon (*COX*2, *ATP*6, *COX*3, *ND*4, *ND4L* and *Cyt*B with ATG; *ND*2, *ATP*8, *ND*6 and *ND*1 with ATT; and *ND*3 with ATA), but two (*COX*1 and *ND*5) use TCG and GTG as the start codon, respectively. Nine PCGs have a complete translation termination codon TAA (*ND*2, *ATP*8, *ATP*6, *ND*3, *ND*5, *ND4L*, *ND*6, *CytB* and *ND*1), whereas the remaining four (*COX*1, *COX*2, *COX*3 and *ND*4) have incomplete stop codon T (Table 2). As a result of comparison of PCGs of 33 *Anopheles* mtgenomes, the start codons of *COX*1 and *ND*5 are highly variable throughout different mtgenomes, *COX*1 and *ND*5 of most species using TCG and GTG as the start codon, respectively. PCGs of all known *Anopheles* mtgenomes use TAA, T or TA as stop codons (Table 4). There are 61 different codons used in *An. lindesayi* mtgenome, and the relative synonymous codon usage (RSCU) values of UUA, UCU, CCU, GCU, CGA and GGA are higher, while those of CUC, CUG, GUC, CCG, UAG, UAA and GGC are relatively lower (Table 5). The usage bias of amino acids for the 13 PCGs was identified in the *An. lingdesayi* mtgenome. Leu has the highest percentage, and Cys has the lowest percentage (Fig. 3).

There are 22 complete tRNAs, with a length range from 64 bp to 72 bp. All tRNAs can fold into the typical clover-leaf structure, containing four stems [double hydrogen uracil (DHU), amino acids (AA), TΨC and anticodons (AC)] and loops [DHU, TΨC, AC and variable (V)] except for tRNA$^{Ser(AGN)}$ which has a large loop instead of the conserved stem-and-loop structure (Fig. 4). There are 18 mismatched base pairs to be found in *An. lindesayi* tRNAs based on the secondary structure analysis with 17 of them being G-U pairs and the remaining one being U-U (Fig. 4). The 16S rRNA is located between tRNA$^{Leu(CUN)}$ and tRNA^{Val}, and 12S rRNA between tRNA^{Val} and CR. The 16S rRNA and 12S rRNA are 1 327 bp and 796 bp in length with the AT contents of 83.12% and 80.28%, respectively.

Table 4 Start/stop codon of 13 protein-coding genes in mtgenomes of 33 *Anopheles* species

Species	*ND*2		*COX*1		*COX*2		*ATP*8		*ATP*6		*COX*3		*ND*3		*ND*5		*ND*4		*ND*4*L*		*ND*6		*CytB*		*ND*1	
	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN	ST	EN
Anopheles subgenus																										
Myzorhynchus Series																										
An. lindesayi	ATT	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATT	TAA
An. atroparvus	ATT	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	ATT	TA	ATG	T	ATG	TAA	ATT	TAA	ATG	T	TTG	TAA
An. quadrimaculatus A	ATT	TAA	TCG	TAA	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATT	TAA
Anopheles Series																										
An. sinensis	ATT	TAA	TCG	TAA	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATT	TAA
Cellia subgenus																										
Neomyzomyia Series																										
An. cracens	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATC	TAA	ATG	TAA	ATG	TAA
An. dirus A	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATG	TAA
An. farauti 1	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATC	TAA	ATG	TAA	ATA	TAA
An. farauti 4	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATC	TAA	ATG	TAA	ATA	TAA
An. hinesorum	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATC	TAA	ATG	TAA	ATA	TAA
An. koliensis	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA
An. punctulatus	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TA	ATG	T	ATG	TAA	ATC	TAA	ATG	T	TTG	TAA
Neocellia Series																										
An. maculatus	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	T	TTG	TAA
Myzomyia Series																										
An. minimus	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	TAA	ATT	TAA
An. culicifacies B	ATC	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	TAA	ATT	TAA
An. culicifacies	ATC	T	ATT	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TA	ATG	T	ATA	TAA	ATC	TAA	ATG	TA	TTG	TAA
Pyretophorus Series																										
An. arabiensis	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	T	TTG	TAA
An. christyi	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATG	TAA	ATC	TAA	ATG	T	TTG	TAA
An. coluzzii	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	T	TTG	TAA
An. epiroticus	ATC	T	ATT	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	T	TTG	TAA
An. gambiae	ATC	TAA	TCG	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	TAA	ATA	TAA
An. melas	ATC	T	ATT	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATA	TAA	ATT	TAA	ATG	T	TTG	TAA
An. merus	ATC	T	ATT	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	T	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	T	TTG	TAA
Nyssorhynchus subgenus																										
Albitarsis Series																										
An. albitarsis F	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
An. albitarsis G	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
An. albitarsis	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	ATG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
An. oryzalimnetes	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
An. deaneorum	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
An. janconnae	ATT	T	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	T	GTG	T	ATG	T	ATG	TAA	ATT	TAA	ATG	T	ATA	TAA
Argyritarsis Series																										
An. darlingi	ATT	TAA	TCG	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA
Kerteszia subgenus																										
An. bellator	ATT	TAA	TCG	T	ATG	T	ATC	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA
An. laneanus	ATT	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA
An. cruzii	ATT	TAA	TCG	T	ATG	T	ATA	TAA	ATG	TAA	ATG	T	ATA	TAA	GTG	TAA	ATG	T	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA
An. homunculus	ATT	TAA	TCG	T	ATG	T	ATT	TAA	ATG	TAA	ATG	TA	ATA	TAA	ATC	TAA	ATG	TAA	ATG	TAA	ATT	TAA	ATG	TAA	ATA	TAA

ST: Start codon; EN: Stop codon.

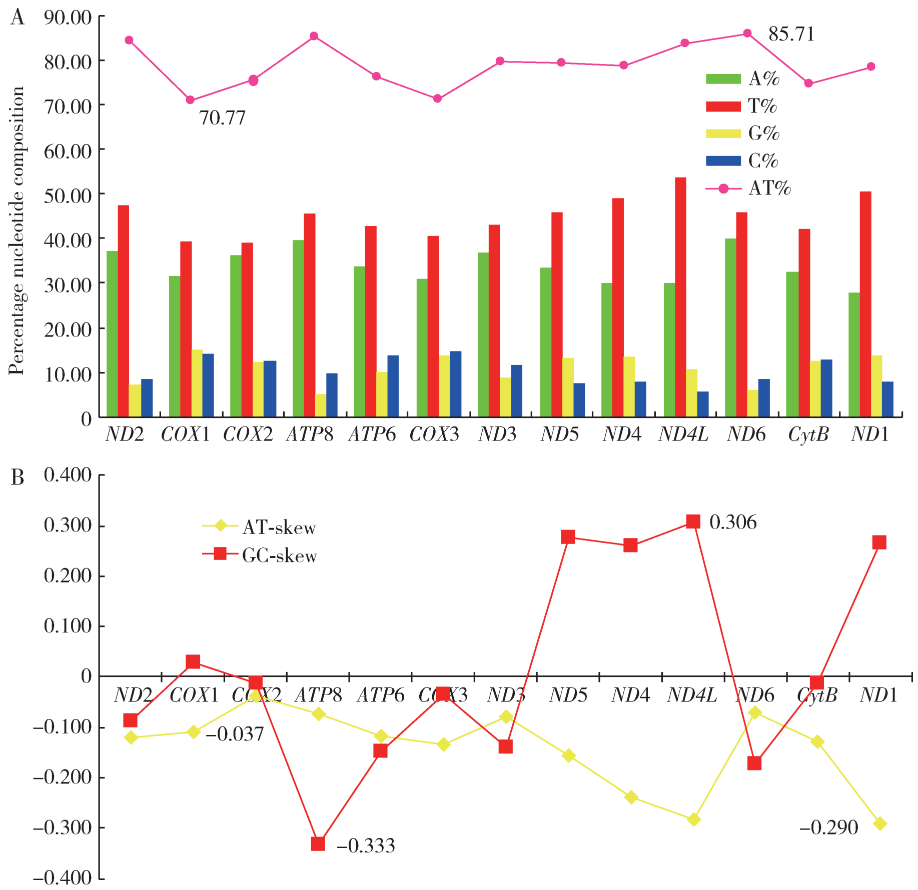

Fig. 2 Nucleotide composition of 13 PCGs in *Anopheles lindesayi* mtgenome

A: Nucleotide composition; B: AT-skew and GC-skew. All values were calculated based on the majority strand.

Table 5 Relative synonymous codon usage (RSCU) in the *Anopheles lindesayi* mtgenome

Codon	RSCU	Codon	RSCU	Codon	RSCU	Codon	RSCU
UUU(F)	1.92	UCU(S)	2.19	UAU(Y)	1.88	UGU(C)	1.86
UUC(F)	0.08	UCC(S)	0.12	UAC(Y)	0.12	UGC(C)	0.14
UUA(L)	5.47	UCA(S)	1.59	UAA(*)	0	UGA(W)	1.90
UUG(L)	0.12	UCG(S)	0.10	UAG(*)	0	UGG(W)	0.10
CUU(L)	0.26	CCU(P)	2.38	CAU(H)	1.73	CGU(R)	0.77
CUC(L)	0.02	CCC(P)	0.22	CAC(H)	0.27	CGC(R)	0.21
CUA(L)	0.12	CCA(P)	1.37	CAA(Q)	1.92	CGA(R)	2.88
CUG(L)	0	CCG(P)	0.03	CAG(Q)	0.08	CGG(R)	0.14
AUU(I)	1.95	ACU(T)	2.01	AAU(N)	1.91	AGU(S)	1.89
AUC(I)	0.05	ACC(T)	0.12	AAC(N)	0.09	AGC(S)	0.19
AUA(M)	1.86	ACA(T)	1.83	AAA(K)	1.68	AGA(S)	1.82
AUG(M)	0.14	ACG(T)	0.04	AAG(K)	0.32	AGG(S)	0.10
GUU(V)	1.83	GCU(A)	2.28	GAU(D)	1.64	GGU(G)	0.69
GUC(V)	0.02	GCC(A)	0.17	GAC(D)	0.36	GGC(G)	0
GUA(V)	2.06	GCA(A)	1.50	GAA(E)	1.93	GGA(G)	2.50
GUG(V)	0.09	GCG(A)	0.05	GAG(E)	0.08	GGG(G)	0.80

*: Stop codon.

The CR is 531 bp in length and located between 12S rRNA and tRNA^{Ile}. It includes the origin site for transcription and replication, and is very rich in A + T content (94.54%). Compared with other 32 *Anopheles* mtgenomes investigated, *An. lindesayi* mtgenome also contains a conservative T-stretch structure, which is located in position 14 979 bp, totally 15 bp in size.

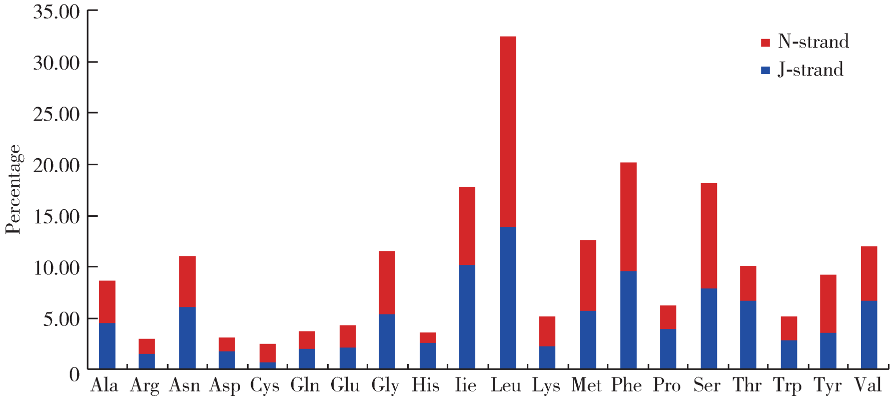

Fig. 3 Percentage of amino acids on both the majority strand (N-strand) and minority strand (J-strand) in *Anopheles lindesayi* mtgenome

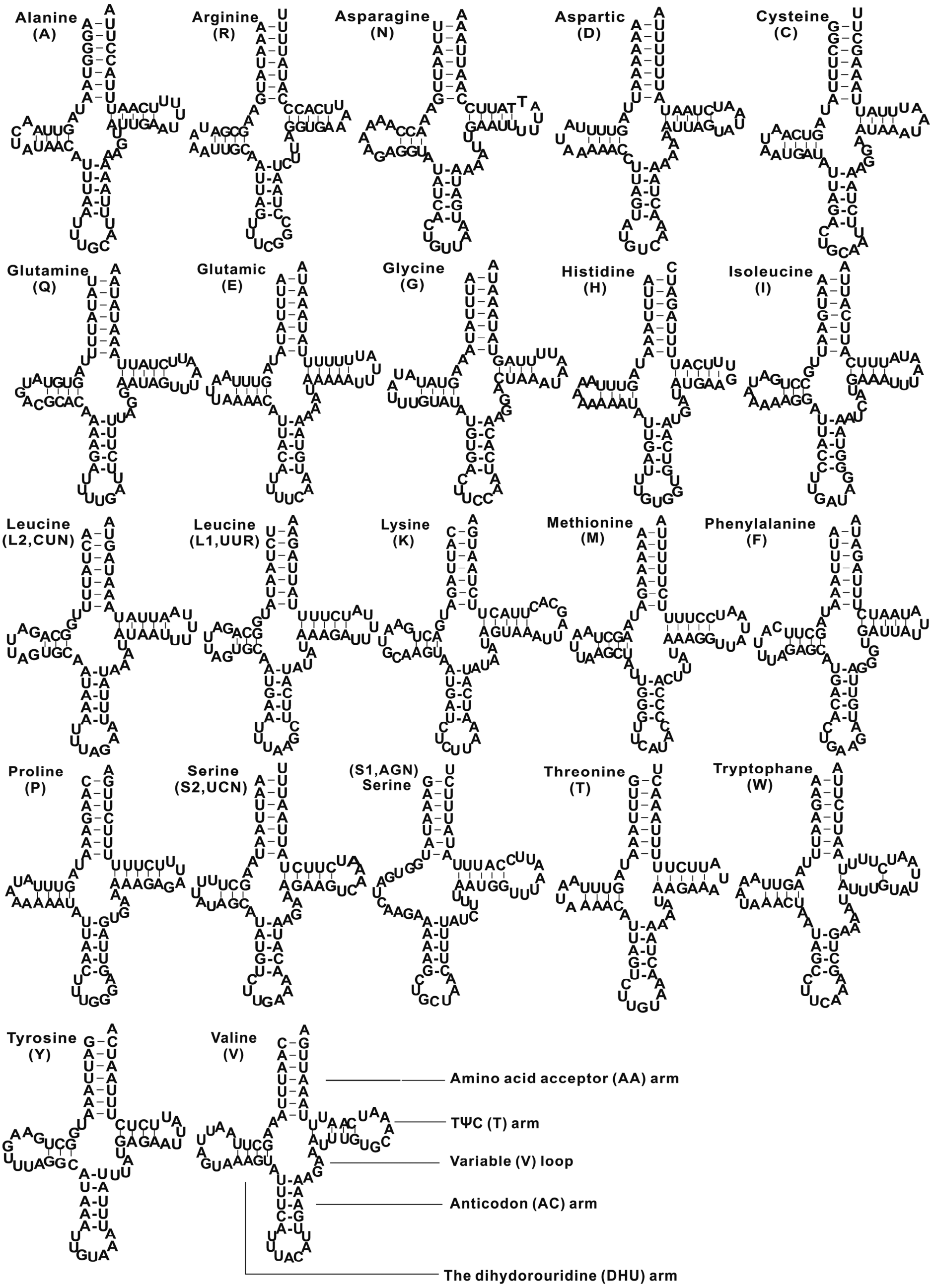

Fig. 4 Inferred secondary structures of 22 tRNAs in *Anopheles lindesayi* mtgenome

The tRNAs are marked with their corresponding amino acids.

3.3 Nucleotide diversity of *Anopheles* mtgenomes (without CR)

A sliding window analysis was conducted to detect regions with high variable evolution rates and intra specific polymorphisms in the known mtgenomes in *Anopheles*. The aligned sequences of four subgenera of *Anopheles* mtgenomes are 14 967 bp (without CR), and the nucleotide diversity (*Pi*) values at nucleotide positions ranged from 0.0142 to 0.18397 (Fig. 5). As a result of comparisons of different types of genes, all of the 13 PCGs had higher *Pi* values ($Pi \geqslant 0.092$). All of the 22 tRNAs had a low *Pi* value ($Pi \leqslant 0.092$), except tRNA$^{\mathrm{His}}$ ($Pi = 0.0991 - 0.1117$), tRNA$^{\mathrm{Thr}}$ ($Pi = 0.0892 - 0.0933$) and tRNA$^{\mathrm{Ser(UCN)}}$ ($Pi = 0.0839 - 0.0953$) that had the higher *Pi* value ($Pi \geqslant 0.092$). The rRNAs had the lowest values [$Pi_{(\mathrm{12S\ rRNA})} = 0.0142 - 0.0605$, $Pi_{(\mathrm{16S\ rRNA})} = 0.0167 - 0.0876$].

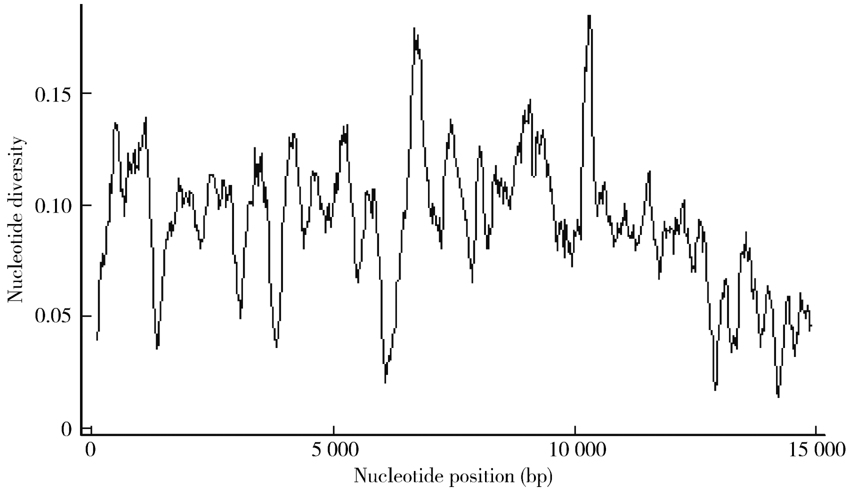

Fig. 5 Nucleotide diversity (*Pi*) of the mtgenomes (without CR) of 33 *Anopheles* species
The nucleotide diversity was calculated using a sliding window analytical procedure (a window of 200 bp with step size of 25 bp was applied in the analysis).

3.4 Phylogenetic analysis of *Anopheles* based on mtgenomes

The phylogenetic analyses based on two methods (ML and BI) and two datasets (PCG and AA) produced identical topologies of the phylogenetic trees for the four subgenera in *Anopheles*, (((*Cellia* + *Anopheles*) + *Nyssorhynchus*, and *Kerteszia*), and the clades for the four subgenera were supported by at least 74% of bootstrap values. The phylogenetic relationships among species, ML-PCG tree (phylogenetic tree based on the concatenated nucleotide sequences of 13 protein-coding genes in mitogenomes of 33 *Anopheles* species inferred by maximum likelihood method), BI-PCG tree (phylogenetic tree based on the concatenated nucleotide sequences of 13 protein-coding genes in mitogenomes of 33 *Anopheles* species inferred by Bayesian inference method), ML-AA tree (phylogenetic tree based on the concatenated amino acid sequences of 13 PCGs in mitogenomes of 33 *Anopheles* species inferred by maximum likelihood method) and BI-AA tree (phylogenetic tree based on the concatenated amino acid sequences of 13 PCGs in mitogenomes of 33 *Anopheles* species inferred by Bayesian inference method) were also identical except for those of some species in the Pyretophorus Series in the subgenus *Cellia* and *An. bellator* and *An. homunculus* in the subgenus *Keterszia* (Figs. 6-9).

In the Pyretophorus, *An. epiroticus* and *An. christyi* were located at the base of the Pyretophorus Series clade in the trees inferred by ML-PCG tree (Fig. 6) and BI-AA tree and ML-AA tree (Figs. 8-9), respectively, whereas *An. epiroticus* + *An. christyi* is jointly linked at the base in the tree from BI-PCG tree (Fig. 7). *An. merus* and *An. Melas* are sister each other based on ML-PCG tree (Fig. 6), while they are not sister based on the BI-PCG tree, BI-AA tree and ML-AA tree analyses (Figs. 7-9). In the subgenus *Keterszia*, *An. bellator* was located at the base of the subgenus *Kerteszia* clade (Fig. 6, Fig. 8) based on ML-PCG tree and ML-AA tree, whereas *An. bellator* + *An. homunculus* was at the base based on the BI-PCG tree and BI-AA tree with 87% and 86% bootstrap values of support (Figs. 8-9).

In subgenus *Anopheles* clade, four species belonging to Anopheles and Myzorhynchus Series were included into the present phylogenetic analyses. *An. atroparvus* and *An. quadrimaculatus*

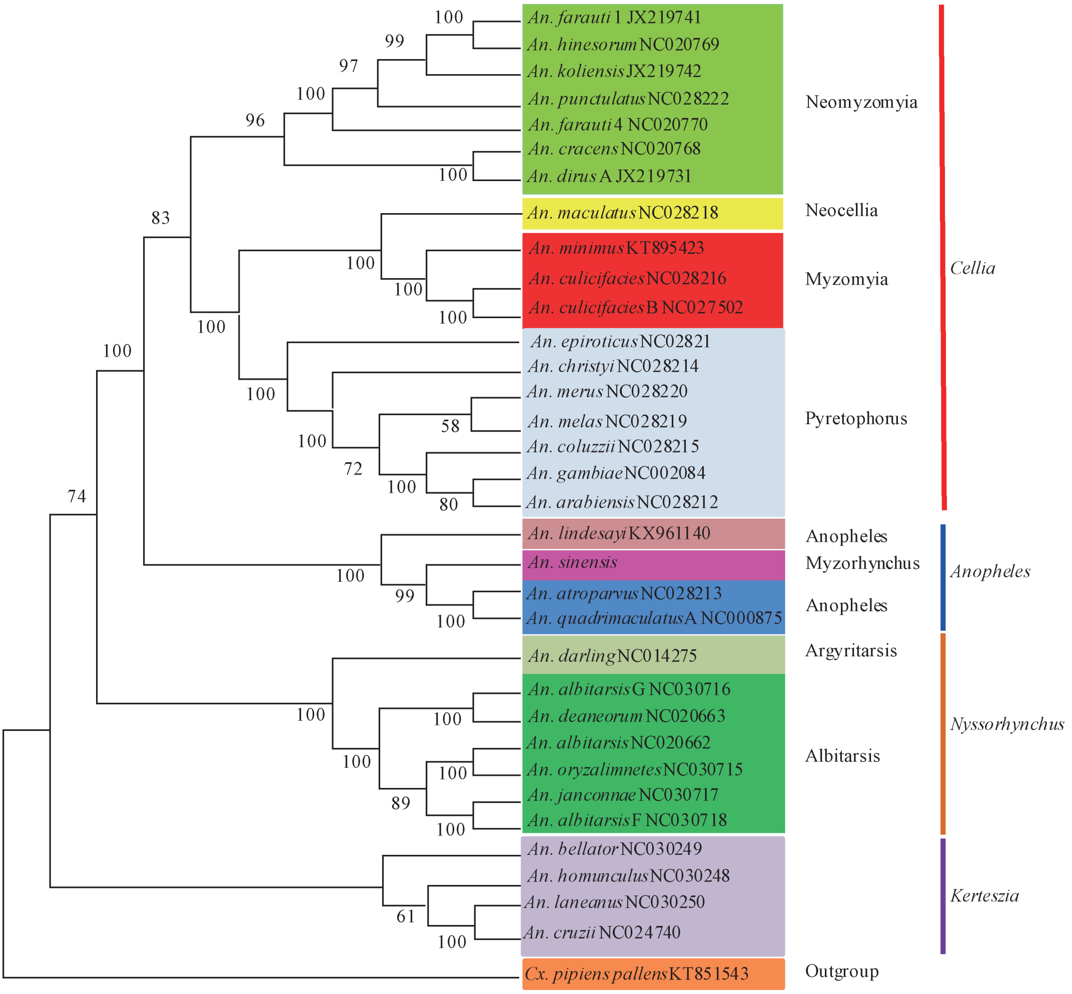

Fig. 6　Phylogenetic tree based on the concatenated nucleotide sequences of 13 protein-coding genes in mtgenomes of 33 *Anopheles* species inferred by maximum likelihood method (ML-PCG tree)

A, belonging to the Maculipennis Group of Anopheles Series, were clustered together as sister each other with 100% bootstrap value of support. *An. lindesayi* belonging to the Lindesayi Group of Anopheles Series, was linked to the base of subgenus *Anopheles*. Strangely, *An. lindesayi* and *An. atroparvus* + *An. quadrimaculatus* A, both belonging to Anopheles Series, were separated by *An. sinensis*, which belongs to the Myzorhynchus Series. The results of phylogenetic analysis were inconsistent with the traditional morphological analysis.

4　DISCUSSION

4.1　Characteristics of *An. lindesayi* mtgenome

The gene arrangement in *An. lindesayi* mtgenome is consistent with the known *Anopheles* mtgenomes (Beard *et al*., 1993; Logue *et al*., 2013; Hua *et al*., 2016; Oliveira *et al*., 2016). The *Anopheles* mtgenomes have higher percentages of A + T, which is consistent with other dipteran species (Zhang *et al*., 2015). In the PCGs of *An. lindesayi* mtgenome, the A + T content of the 3rd codon position is obviously higher than those of the 1st and 2nd. As in the known mtgenomes of other *Anopheles* species (Kim *et al*., 2014), the codon of *An. lindesayi* mtgenome prefers to use A and T in the 3rd position, which provides further evidence for high background mutational pressure toward AT nucleotides at the 3rd position.

The *An. lindesayi* anticodons are consistent with those in other known *Anopheles* mtgenomes (Krzywinski *et al*., 2011; Logue *et al*., 2013). All tRNAs form the typical clover-leaf structure, but the DHU arm of $\mathrm{tRNA}^{\mathrm{Ser(AGN)}}$ in the *An. lindesayi* mtgenome is a large loop instead of the conserved stem-and-loop structure, which is a typical feature of metazoan mtgenomes (Negrisolo *et al*., 2011). The secondary structures of all tRNAs of mosquito mtgenomes can form weak bonds that affect the thermodynamic stability and non-canonical pairs (Varani and Mcclain, 2000; Gutell *et al*., 2002). There are 18 mismatched base pairs to be identified in *An. lindesayi* mtgenome, and the mismatch is a

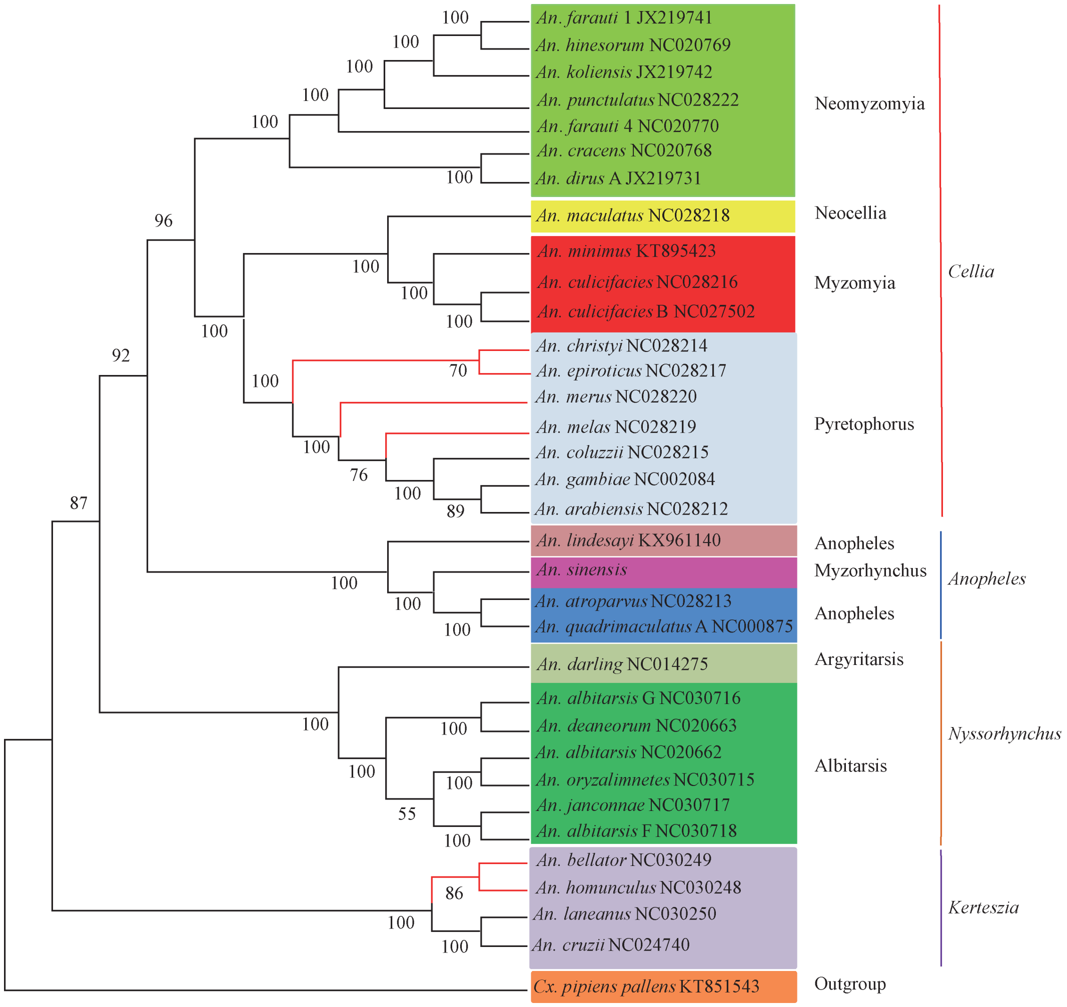

Fig. 7 Phylogenetic tree based on the concatenated nucleotide sequences of 13 protein-coding genes in mtgenomes of 33 *Anopheles* species inferred by Bayesian inference method (BI-PCG tree)

Branches in red indicate the difference between Figs. 6 and 7.

common phenomenon but could be corrected in posttranscriptional RNA editing processes (Masta and Boore, 2004). The position of the 16S rRNA and 12S rRNA in the *An. lindesayi* mtgenome both correspond to those of the mtgenomes in other known *Anopheles* species (Krzywinski *et al*., 2011; Hua *et al*., 2015).

4.2 Comparison of 33 *Anopheles* mtgenomes

In 33 *Anopheles* mtgenomes investigated, most PCGs have a start codon ATN, such as the genes *ND*2, *COX*2, *ATP*8, *ATP*6, *COX*3, *ND*3, *ND*4, *ND4L*, *ND*6 and *CytB*, whereas the other PCGs contain some unusual start codons, such as GTG, TCG and TTG. The use of the start codons GTG and TCG has been documented in various organisms, including *Anopheles* species (Beard *et al*., 1993; Krzywinski *et al*., 2006). The PCGs in *Anopheles* mtgenomes all use TAA, T or TA as stop codon, and the variation of stop codons seems less than that of start codons. The presence of incomplete stop codons is common in metazoan mtgenomes, and these incomplete stop codons could be completed via post-transcriptional polyadenylation (Ojala *et al*., 1981).

In 33 *Anopheles* mtgenomes, the CRs, due to large differences in length, are difficult to discuss their phylogenetic relationship. The tRNAs and rRNAs, all showing quite low nucleotide diversity from the sliding window analysis, do not seem to be good markers for taxonomy and evolutionary studies. The PCGs, showing higher nucleotide diversity, would be the most suitable markers to elucidate the phylogenetic relationships at the subgenus and genus levels. All of these conclusions fit the known mtgenomes of Culicinae mosquitoes (Demari-Silva *et al*., 2015; Luo *et al*., 2016).

4.3 Phylogeny of *Anopheles* based on mtgenomes

Our results strongly supported each of the subgenera *Cellia*, *Anopheles*, *Nyssorhynchus* and *Kerteszia* to be a monophyletic group with at least 74% bootstrap values, which is consistent with previous phylogenetic analyses based on morphological and molecular data (Sallum *et al*., 2000, 2002;

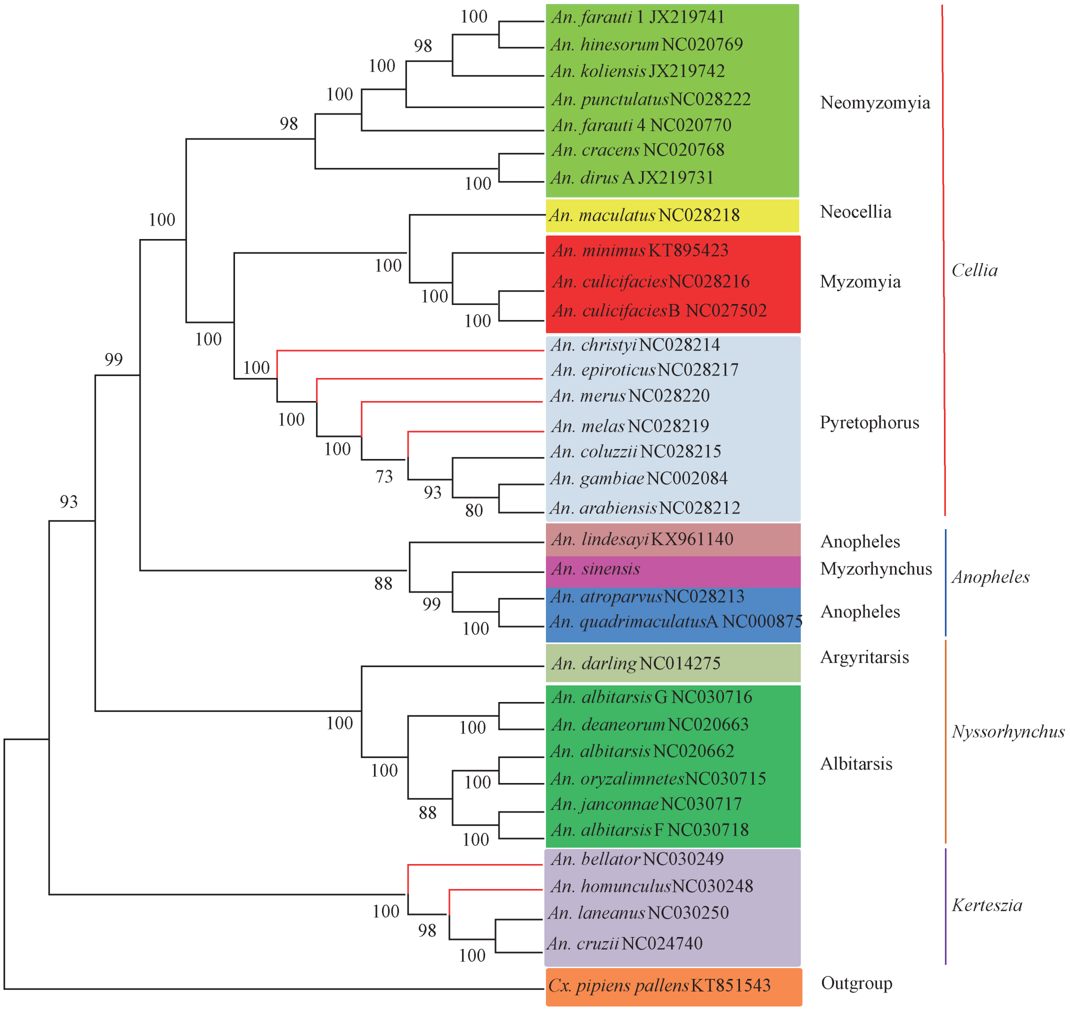

Fig. 8 Phylogenetic tree based on the concatenated amino acid sequences of 13 protein-coding genes in mtgenomes of 33 *Anopheles* species inferred by maximum likelihood method (ML-AA tree)
Branches in red indicate the difference between Figs. 6 and 8.

Krzywinski *et al.*, 2001; Harbach and Kitching, 2005, 2016; Mohanty *et al.*, 2009; Wang *et al.*, 2014). The subgenera *Cellia* and *Anopheles* were clustered together as sister each other with at least 92% bootstrap value of support, and the subgenus *Kerteszia* was recognized as the earliest-derived clade in the overall genus-level phylogenetic relationship explored in the present study, [*Kerteszia* + ((*Cellia* + *Anopheles*) + *Nyssorhynchus*)], which is consistent with previous studies (Foley *et al.*, 1998; Harbach and Kitching, 2005, 2010; Harbach, 2007).

The phylogenetic uncertainty part are four species in the Pyretophorus Series in the subgenus *Cellia* and the *An. bellator* + *An. homunculus* clade in the subgenus *Keterszia*, might stem from the different phylogenetic construction methods. The phylogenetic position in consistence of two species in the subgenus *Anopheles*, compared with the traditional morphological analysis, also needs to be further elucidated with more molecular data and detailed morphological studies.

The resolution of reliable phylogenetic relationships is the base for the establishment of natural taxonomic system. *Anopheles* mosquitoes are overwhelmingly important due to transmission of malarial and filarial parasites to humans. Compared with the whole *Anopheles* species (> 400), the phylogeny of 33 *Anopheles* species is not enough to draw the outline of the phylogenetic relationships of the genus. Therefore, more studies with a variety of species are needed to provide further insights into the relationships in the genus *Anopheles*.

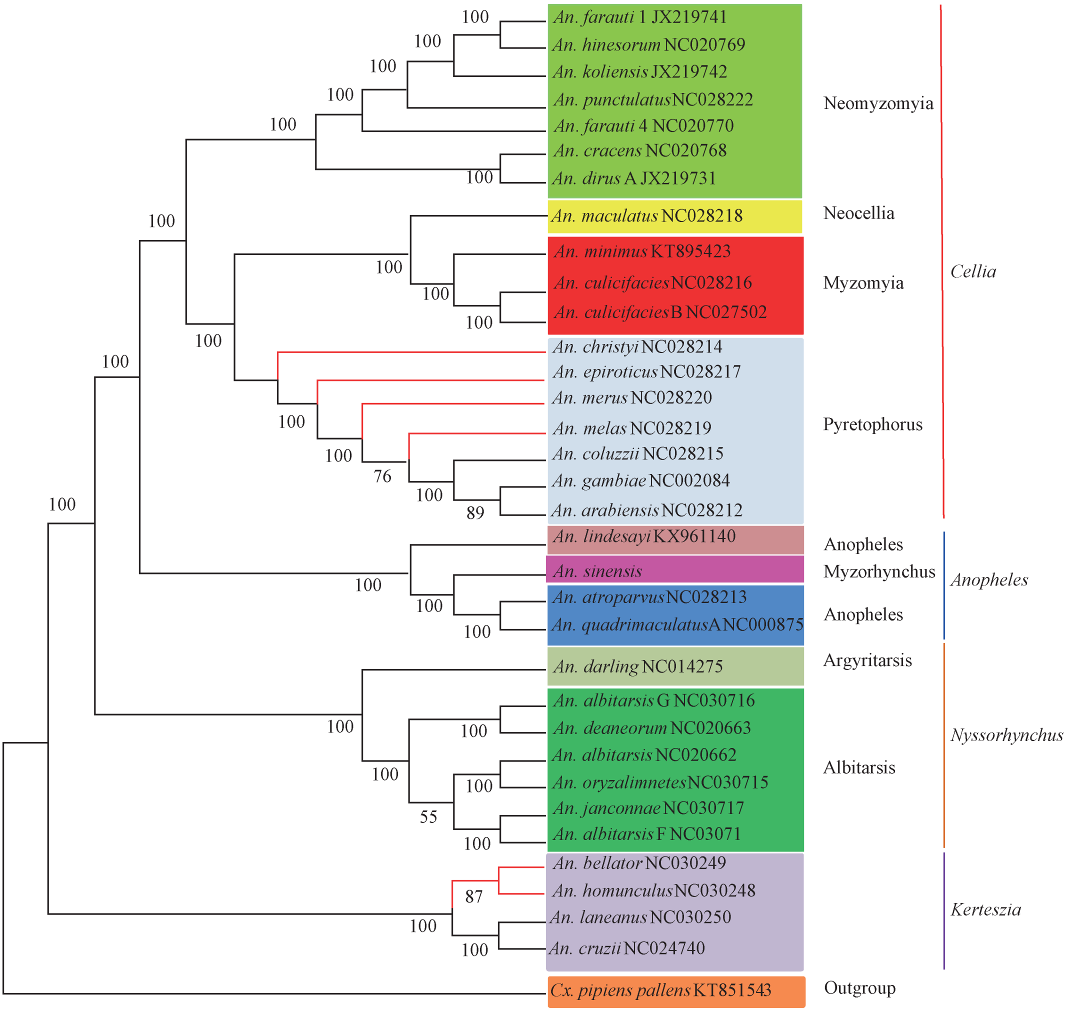

Fig. 9 Phylogenetic tree based on the concatenated amino acid sequences of 13 protein-coding genes in mtgenomes of 33 *Anopheles* species inferred by Bayesian inference method (BI-AA tree)

Branches in red indicate the difference between Figs. 6 and 9.

References

Alzohairy AM, 2011. BioEdit: an important software for molecular biology. *Gerf. Bull. Biosci.*, 2(1): 60 – 61.

Beard CB, Hamm DM, Collins FH, 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.*, 2(2): 103 – 124.

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Res.*, 27 (8): 1767 – 1780.

Boore JL, 2006. The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol. Evol.*, 21(8): 439 – 446.

Boore JL, Macey JR, Medina M, 2005. Sequencing and comparing whole mitochondrial genomes of animals. *Methods Enzymol.*, 395: 311 – 348.

Cameron SL, 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annu. Rev. Entomol.*, 59: 95 – 117.

Demari-Silva B, Foster PG, Oliveira TMPD, Bergo ES, Sanabani SS, Pessoa R, Sallum MAM, 2015. Mitochondrial genomes and comparative analyses of *Culex camposi*, *Culex coronator*, *Culex usquatus*, and *Culex usquatissimus*, (Diptera: Culicidae), members of the coronator group. *BMC Genomics*, 16: 831.

Foley DH, Bryan JH, Yeates D, Saul A, 1998. Evolution and systematics of *Anopheles*: insights from a molecular phylogeny of Australasian mosquitoes. *Mol. Phylogenet. Evol.*, 9(2): 262 – 275.

Glick JI, 1992. Illustrated key to the female *Anopheles* of southwestern Asia and Egypt (Diptera: Culicidae). *Mosq. Syst.*, 24(2): 125 – 153.

Gutell RR, Lee JC, Cannone JJ, 2002. The accuracy of ribosomal RNA comparative structure models. *Curr. Opin. Struct. Biol.*, 12(3): 301 – 310.

Hao YJ, Zou YL, Ding YR, Xu WY, Yan ZT, Li XD, Fu WB, Li TJ, Chen B, 2017. Complete mitochondrial genomes of *Anopheles stephensi* and *An. dirus* and comparative evolutionary mitochondriomics of 50 mosquitoes. *Sci. Rep.*, 7(1): 7666 – 7679.

Harbach RE, 2007. The Culicidae (Diptera): a review of taxonomy, classification and phylogeny. *Zootaxa*, 1668: 591 – 638.

Harbach RE, Kitching IJ, 2005. Reconsideration of anopheline mosquito phylogeny (Diptera: Culicidae: Anophelinae) based on morphological data. *Syst. Biodivers.*, 3(4): 345 – 374.

Harbach RE, Kitching IJ, 2010. Phylogeny and classification of the Culicidae (Diptera). *Syst. Entomol.*, 23(4): 327 – 370.

Harbach RE, Kitching IJ, 2016. The phylogeny of Anophelinae revisited: inferences about the origin and classification of *Anopheles* (Diptera: Culicidae). *Zool. Scr.*, 45(1): 34 – 47.

Harrison BA, Rattanarithikul R, Peyton EL, Mongkolpanya K, 1990. Taxonomic changes, revised occurrence records and notes on the Culicidae of Thailand and neighboring countries. *Mosq. Syst.*, 22 (3): 196 – 227.

Hua YQ, Ding YR, Yan ZT, Si FL, Luo QC, Chen B, 2016. The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Sci.*, 23(3): 353-365.

Hua YQ, Yan ZT, Fu WB, He QY, Zhou Y, Chen B, 2015. Sequencing and analysis of the complete mitogenome in *Anopheles culicifacies* species B (Diptera: Culicidae). *Mitochondr. DNA Part A*, 27(4): 2909-2910.

Kazutaka K, Standley DM, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, 30(4): 772-780.

Kim H, Sames WS, Lee I, Lee D, Kim H, Rueda L, Klein T, 2009. Overwintering of *Anopheles lindesayi japonicus* larvae in the Republic of Korea. *J. Am. Mosq. Control Assoc.*, 25(1): 32-37.

Kim MJ, Wang AR, Park JS, Kim I, 2014. Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. *Gene*, 549(1): 97-112.

Krzywinski J, Grushko OG, Besansky NJ, 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.*, 39(2): 417-423.

Krzywinski J, Li C, Morris M, Conn JE, Lima JB, Povoa MM, Wilkerson RC, 2011. Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex. *Mol. Phylogenet. Evol.*, 58(3): 469-477.

Krzywinski J, Wilkerson RC, Besansky NJ, 2001. Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. *Mol. Phylogenet. Evol.*, 18(3): 479-487.

Librado P, Rozas J, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11): 1451-1452.

Logue K, Chan ER, Phipps T, Small ST, Reimer L, Henryhalldin C, Sattabongkot J, Siba PM, Zimmerman PA, Serre D, 2013. Mitochondrial genome sequences reveal deep divergences among *Anopheles punctulatus* sibling species in Papua New Guinea. *Malar. J.*, 12(1): 64-74.

Lowe TM, Eddy SR, 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, 25(5): 955-964.

Luo QC, Hao YJ, Meng F, Li TJ, Ding YR, Hua YQ, Chen B, 2016. The mitochondrial genomes of *Culex tritaeniorhynchus* and *Culex pipiens pallens* (Diptera: Culicidae) and comparison analysis with two other *Culex* species. *Parasit. Vectors*, 9(1): 406-416.

Ma Y, Ma YJ, Lin L, Wang Y, 2013. Phylogenetic relationship among some species of genus *Anopheles* subgenus *Anopheles* (Diptera: Culicidae) in China: based on rDNA-ITS2 sequences. *Chin. J. Vector Biol. Control*, 24(5): 382-388. [马颖，马雅军，林琳，王琰，2013. 基于 rDNA-ITS2 序列的中国按蚊属按蚊亚属部分种类的系统发育关系研究. 中国媒介生物学及控制杂志，24(5): 382-388]

Masta SE, Boore JL, 2004. The complete mitochondrial genome sequence of the spider *Habronattus oregonensis* reveals rearranged and extremely truncated tRNAs. *Mol. Biol. Evol.*, 21(5): 893-902.

Masta SE, Boore JL, 2008. Parallel evolution of truncated transfer RNA genes in arachnid mitochondrial genomes. *Mol. Biol. Evol.*, 25(5): 949-959.

Mitchell SE, Cockburn AF, Seawright JA, 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*, 36(6): 1058-1073.

Mohanty A, Swain S, Kar SK, Hazra RK, 2009. Analysis of the phylogenetic relationship of *Anopheles* species, subgenus *Cellia* (Diptera: Culicidae) and using it to define the relationship of morphologically similar species. *Infect. Genet. Evol.*, 9(6): 1204-1224.

Moreno M, Marinotti O, Krzywinski J, Tadei WP, James AA, Achee N, Conn JE, 2010. Complete mtDNA genomes of *Anopheles darlingi*, and an approach to anopheline divergence time. *Malaria J.*, 9(1): 127-127.

Negrisolo, E, Babbucci M, Patarnello T, 2011. The mitochondrial genome of the ascalaphid owlfly *Libelloides macaronius* and comparative evolutionary mitochondriomics of neuropterid insects. *BMC Genomics*, 12(1): 221-246.

Ojala D, Montoya J, Attardi G, 1981. TRNA punctuation model of RNA processing in human mitochondrial. *Nature*, 290(5806): 470-474.

Oliveira TMP, Foster PG, Bergo ES, Nagaki SS, Sanabani SS, Marinotti O, Marinotti PN, Sallum MAM, 2016. Mitochondrial genomes of *Anopheles* (*Kerteszia*) (Diptera: Culicidae) from the Atlantic forest, Brazil. *J. Med. Entomol.*, 53(4): tjw001.

Perna NT, Kocher TD, 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.*, 41(3): 353-358.

Posada D, Crandall KA, 1998. MODELTEST, Ver. 3.06: testing the model of DNA substitution. *Bioinformatics*, 14(9): 817-818.

Reid JA, 1968. Anopheline mosquitoes of Malaya and Borneo. Studies from the Institute for Medical Research, Malaysia.

Ronquist F, Huelsenbeck JP, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12): 1572-1574.

Sallum MAM, Schultz TR, Foster PG, Aronstein K, Wirtz RA, Wilkerson RC, 2002. Phylogeny of Anophelinae (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. *Syst. Entomol.*, 27(3): 361-382.

Sallum MAM, Schultz TR, Wilkerson RC, 2000. Phylogeny of Anophelinae (Diptera Culicidae) based on morphological characters. *Ann. Entomol. Soc. Am.*, 93(4): 745-775.

Sames WJ, Kim HC, Chong ST, Harrison BA, Lee WJ, Rueda LM, Klein TA, 2008. *Anopheles lindesayi japonicus* Yamada (Diptera: Culicidae) in Korea: comprehensive review, new collection records, and description of larval habitats. *J. Vector Ecol.*, 33(1): 99-106.

Stamatakis A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9): 1312-1313.

Talavera G, Castresana J, 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.*, 56(4): 564-577.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2014. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.*, 28(10): 2731-2739.

Varani G, Mcclain WH, 2000. The G x U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.*, 1(1): 18-23.

Wang G, Li C, Guo X, Xing D, Dong Y, Wang Z, Zhang Y, Liu M, Zheng Z, Zhang H, 2012. Identifying the main mosquito species in China based on DNA barcoding. *PLoS ONE*, 7(10): e47051.

Wang G, Li C, Guo X, Xing D, Dong Y, Zhao T, 2014. Molecular phylogenetic analysis of the subgenera *Anopheles* and *Cellia* (Diptera: Culicidae) based on nuclear ribosomal sequences. *Afr. Entomol.*, 22(3): 660-669.

Wei SJ, Shi M, Sharkey MJ, Cornelis VA, Chen XX, 2010. Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to holometabolous insects. *BMC Genomics*, 11(1): 371-386.

Zhang NX, Yu G, Li TJ, He QY, Zhou Y, Si FL, Ren S, Chen B, 2015. The complete mitochondrial genome of *Delia antiqua* and its implications in dipteran phylogenetics. *PLoS ONE*, 10(10): e0139736.

Zhang NX, Zhang YJ, Yu G, Chen B, 2013. Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. *Acta Entomol. Sin.*, 56(4): 398-407. [张乃心，张玉娟，余果，陈斌，2013. 双翅目昆虫

线粒体基因组结构特点及其测序通用引物的设计和应用. 昆虫学报, 56(4): 398-407]

Zhao L, Zheng ZM, Huang Y, Sun HM, 2010. A comparative analysis of mitochondrial genomes in Orthoptera (Arthropoda: Insecta) and genome descriptions of three grasshopper species. *Zool. Sci.*, 27(8): 662-672.

林氏按蚊线粒体全基因组序列的测定及基于线粒体基因组的按蚊属系统发育分析

毛启萌, 李廷景, 付文博, 闫振天, 陈　斌*

(重庆师范大学昆虫与分子生物学研究所, 重庆市媒介昆虫重点实验室, 重庆 401331)

摘要:【目的】对林氏按蚊 *Anopheles lindesayi* 完整的线粒体基因组进行测序及分析,依据已知的线粒体基因组构建并讨论按蚊属蚊虫的分子系统发育关系。【方法】对林氏按蚊线粒体基因组进行测序、注释,并对其基本特征和基本组成进行分析。基于串联的 13 个蛋白质编码基因的核苷酸序列和氨基酸序列,用 ML 法和贝叶斯法构建林氏按蚊和按蚊属其他 32 种蚊虫的系统发育树,据此探讨按蚊属蚊虫的系统发育关系和系统分类。【结果】林氏按蚊线粒体基因组全长为 15 366 bp,包含 13 个蛋白质编码基因,22 个 tRNA 基因,2 个 rRNA 基因和一段控制区。林氏按蚊线粒体基因组呈现明显的 AT 偏斜和 GC 偏斜,AT 偏斜为正,GC 偏斜为负。除了 *COX*1 使用 TCG 和 *ND*5 使用 GTG 作为起始密码子以外,其他蛋白质编码基因的起始密码子均遵循 ATN 原则;终止密码子为 TAA 或者 T。除了 $\text{tRNA}^{\text{Ser(AGN)}}$ 以外,其他的 tRNA 基因均呈现典型的三叶草二级结构。控制区 AT 含量最高,为 94.54%。滑窗分析显示蛋白质编码基因是用于构建亚属或属水平系统发育关系的最佳分子标记。系统发育树强烈支持塞蚊亚属 *Cellia*、按蚊亚属 *Anopheles*、徕蚊亚属 *Nyssorhynchus* 和柯特蚊亚属 *Kerteszia* 均为单系群。小五斑按蚊 *An. atroparvus* 和四斑按蚊 *An. quadrimaculatus* A 这两个种聚到一起,从传统的形态分类上讲,它们和林氏按蚊均属于按蚊亚属按蚊系蚊虫。但本研究构建的 4 个系统发育树均显示,(小五斑按蚊 *An. atroparvus* + 四斑按蚊 *An. quadrimaculatus* A)和林氏按蚊被属于迈蚊系的中华按蚊分开,这为两个系的分类提供了新的论点。【结论】本研究获得了林氏按蚊的完整的线粒体基因组,探析了按蚊属的线粒体基因组特征和系统发育关系,为进一步研究蚊科线粒体基因组和系统发育关系提供了依据。

关键词: 蚊科; 按蚊属; 林氏按蚊; 线粒体基因组; 核苷酸组成; 系统发育

中图分类号: Q969　　**文献标识码:** A　　**文章编号:** 0454-6296(2019)01-0101-16

(责任编辑: 马丽萍)